心理学报 2022, Vol. 54, No. 8, 996−1008
Acta Psychologica Sinica

https://doi.org/10.3724/SP.J.1041.2022.00996

一种简单有效的 *Q* 矩阵修正新方法*

李 佳 毛秀珍 韦 嘉

(四川师范大学教育科学学院, 成都 610066)

摘 要 *Q* 矩阵的正确性是影响题目参数估计和被试分类准确性的重要因素。针对 *Q* 矩阵修正问题, 首先提出了一种简单有效的新方法(ORDP)。然后, 模拟研究通过改变被试知识状态的分布、样本容量(*N*)、测验长度(*L*)、*Q* 矩阵错误率(*M*)、项目质量(*Iq*)和属性层级结构, 比较了 ORDP 与已有方法(R、RMSEA 和 HD)的表现。研究表明:(1) 当知识状态服从均匀分布时, ORDP 方法在所有层级结构下最优; 当知识状态服从多元正态分布时, RMSEA 和 ORDP 表现没有明显差异, 除独立结构外, RMSEA 方法均稍优于 ORDP 方法; (2) 各方法在多元正态分布下的修正效果不及均匀分布时的修正结果; (3) *N*、*L*、*M*、*Iq* 和属性层级结构对 4 种方法的表现均有明显影响; (4) 基于 Tatsuoka (1984) 分数减法数据的修正结果表明, 采用 ORDP 方法修正的 *Q* 矩阵与数据拟合最优。

关键词 认知诊断, *Q* 矩阵修正, ORDP 方法, DINA 模型

分类号 B841

1 引言

2020 年 6 月, 中共中央国务院发布的《深化新时代教育评价改革总体方法》中明确提出“改革学生评价, 创新德智体美劳过程性评价办法”的发展目标。可见, 教育评价越来越强调过程性评价。认知诊断理论(cognitive diagnostic theory, CDT)运用认知心理学知识分析考生的认知过程、加工技能和知识结构, 并结合现代测量学知识进行诊断分析, 能够提供细粒度、多维度的评估结果, 适应“过程性评价”的要求, 具有重要研究与实践价值。

Q 矩阵表征了项目与属性的关系, 是 CDT 的基础, 也体现了 CDT 与项目反应理论(item response theory, IRT)和经典测量理论(classical test theory, CTT)的不同。事实上, *Q* 矩阵通常由领域专家标定或通过参数估计而得。前者不仅容易受到主观因素的影响, 还大大增加了测验开发的成本和专家工作量; 后者仅仅依据数据分析, 又缺少专家对项目特征的分析, 往往不符合实际情况。有研究表明, 即使 *Q* 矩阵的元素存在少量错误也会增大参数估计误差并降低被试诊断正确率(Rupp & Templin, 2008; de la Torre, 2009; 涂冬波 等, 2012)。*Q* 矩阵标定的准确性和复杂性影响着认知诊断评估在实践中的应用和发展(DeCarlo, 2011)。于是, 检验 *Q* 矩阵的正确性, 并对其进行修正具有重要意义。

针对 *Q* 矩阵估计或修正问题, 研究者们从不同视角提出了多种方法。例如, 基于最优项目区分度视角提出了 δ 法(de la Torre, 2008)、γ 法(涂冬波 等, 2012)和 ζ^2 法(de la Torre & Chiu, 2016)等。它们的核心思想是选择具有最优项目区分度或属性区分度的属性模式作为项目 q 向量。这类方法简单易懂, 计算也比较简便。其中, δ 方法只考虑了项目区分两个极端被试组的能力, 不能反映全体被试的信息; ζ^2 方法虽反映全体被试的信息, 但对样本量的要求较高(汪大勋 等, 2019); γ 方法提出了先筛选再修正的思路, 但容易漏掉参数合理但有误的项目。又如, 基于参数估计视角提出了极大似然估计(maximum likelihood estimation, MLE)、边际极大似然估计(marginal maximum likelihood estimation, MMLE) (Wang, Song, et al., 2018)和贝叶斯估计

收稿日期: 2021-06-30

* 重庆市教育科学“十三五”规划课题基金(2020-GX-401)的资助。

通信作者: 毛秀珍, E-mail: maomao_wanli@163.com

(Chung, 2019; Chen et al., 2018; DeCarlo, 2012)方法。它们都是常用的参数估计方法。其中, MLE 和 MMLE 采用 EM 算法对 Q 矩阵进行重复修正, 修正率较高但比较耗时, 而基于贝叶斯的方法过程复杂且易受先验分布的影响。再如, 从模型与数据的绝对拟合视角提出了 S 统计量(Liu et al., 2012)和残差(Chen, 2017)等多种方法。其中, S 统计量方法表达了正确作答项目与项目对的观察概率分布和预测概率分布的欧氏距离, 而残差方法基于项目对的观察反应和预测反应构建了相关或对数比的残差, 但它们的计算均较为繁琐。

综上, 上述方法各有优势与不足。特别地, 大部分模型数据拟合方法都视观察反应(概率)分布和预测反应(概率)分布为两个独立的分布, 通过建构反映二者的一致性或差异性指标来修正 Q 矩阵。本研究基于观察反应和预测反应将作答反应分为 4 个类别, 构建了一种简单高效且适用于简化模型和饱和模型的 Q 矩阵修正方法。然后, 开展 Monte Carlo 模拟实验, 在多种实验条件下比较新方法与近似误差均方根(root mean square error of approximation, RMSEA) (Kang et al., 2019)、残差指标(residual-based statistic, R) (Yu & Cheng, 2020)和海明距离(hamming distance, HD)方法(汪大勋 等, 2018)在 Q 矩阵修正中的表现。最后, 以 Tatsuoka (1984)的分数减法数据为例, 考察各方法对专家标定的 Q 矩阵的修正情况和修正后模型数据的拟合情况。

为行文方便, 下文以 i 、 j 和 z 分别表示被试、项目和项目可能的得分值。N、K、L、M 和 Iq 分别表示被试人数、测验考察的属性个数、测验长度、Q 矩阵错误率和项目质量。$\boldsymbol{\alpha}_l(l=1,2,\cdots,2^K)$ 和 $\boldsymbol{q}_{jc}(c=1,2,\cdots,2^K-1)$ 分别表示可能的知识状态(knowledge states, KS)和属性考察模式。 y_{ij} 与 η_{ij} 分别表示被试 i 在项目 j 上的观察反应和理想反应。论文第二部分介绍了新方法、R、RMSEA 和 HD 方法, 第三和第四部分分别是模拟数据和实测数据的研究设计与结果分析, 第五部分是结论和讨论。

2　Q 矩阵修正方法

Q 矩阵与模型数据拟合有着密切联系。理论上, 正确的 q 向量应该使模型数据拟合最优。基于绝对拟合指标或相对拟合指标可以判断模型与数据的拟合程度。其中, 绝对拟合指标的方法的核心在于构建反映观察反应和预测反应的差异性或一致性指标。本研究结合观察反应和预测反应将作答反应细分为 4 个类别, 并根据各个类别预测人数比例分布提出一种基于模型数据拟合视角的 Q 矩阵修正方法: 最优反应分布纯度(optimization of response distribution purity, ORDP)方法。

2.1　ORDP 方法

经典决策树是基于某种划分准则, 不断将数据集划分为纯度更高, 不确定性更小的子集的算法。而基尼系数作为经典决策树中最优特征的选择指标, 表示从数据集中随机抽取的两个样本所属类别不一样的概率, 反映了数据集的纯度。其值越小, 数据集的纯度越高。例如, 假设离散型随机变量所有可能的取值为 $h(h=1,2,\cdots,H)$, 对应的概率记为 P_h, 则基尼系数为: $Gini=\sum_{h=1}^{H}P_h(1-P_h)=1-\sum_{h=1}^{H}P_h^2$ 。

根据被试在项目 j 上的反应, 可令 KS 为 $\boldsymbol{\alpha}_l$ $(l=1,2,\cdots,2^K)$ 的总人数和 $\boldsymbol{\alpha}_l$ 中答对项目 j 的人数分别为 N_l 和 r_{lj} 。再根据认知诊断模型(cognitive diagnosis model, CDM), 可得 KS 为 $\boldsymbol{\alpha}_l$ 的被试预测正确作答项目 j 的概率为 $P(y_j=1|\boldsymbol{\alpha}_l)$。那么, 在 r_{lj} 名观察反应为 1 的被试中预测有 $r_{lj}\cdot P(y_j=1|\boldsymbol{\alpha}_l)$ 名被试正确作答和 $r_{lj}\cdot(1-P(y_j=1|\boldsymbol{\alpha}_l))$ 名被试错误作答。它们的人数比例分别为 $f_{lj(1,1)}=r_{lj}\cdot P(y_j=1|\boldsymbol{\alpha}_l)/N_l$ 和 $f_{lj(1,0)}=r_{lj}\cdot(1-P(y_j=1|\boldsymbol{\alpha}_l))/N_l$ 。同理, 在 N_l-r_{lj} 名观察反应为 0 的被试中预测有 $(N_l-r_{lj})\cdot P(y_j=1|\boldsymbol{\alpha}_l)$ 名被试正确作答和 $(N_l-r_{lj})\cdot(1-P(y_j=1|\boldsymbol{\alpha}_l))$ 名被试错误作答。它们的人数比例分别为 $f_{lj(0,1)}=(N_l-r_{lj})\cdot P(y_j=1|\boldsymbol{\alpha}_l)/N_l$ 和 $f_{lj(0,0)}=(N_l-r_{lj})\cdot(1-P(y_j=1|\boldsymbol{\alpha}_l))/N_l$ 。于是, 结合 $\boldsymbol{\alpha}_l$ 类被试在项目 j 上的观察反应和预测反应, 可以将反应分为 $(O1,E1)$ 、$(O1,E0)$ 、$(O0,E1)$ 和 $(O0,E0)$ 四个类别, 其人数比例分布如表 1 所示。例如, $(O1,E1)$ 表示 $\boldsymbol{\alpha}_l$ 类被试在项目 j 上观察反应和预测反应均为 1 的被试反应类别。值得注意的是, 预测反应不同于理想反应。前者指依据 CDM 计算而得的每个被试预测反应为 1 和 0 的可能性。后者指在无失误无猜测的条件下, 当被试掌握了正确作答项目所要求的属性时, 则其理想反应为 1, 否则为 0。依据观察反应和预测反应, 每一类 KS 的被试就被分到如表 1 所示的 4 种反应类别。

表 1　$\boldsymbol{\alpha}_l$ 类被试在项目 j 上 4 种反应类别的人数比例分布

反应类别	$(O1, E1)$	$(O1, E0)$	$(O0, E1)$	$(O0, E0)$
期望人数比例	$f_{lj(1,1)}$	$f_{lj(1,0)}$	$f_{lj(0,1)}$	$f_{lj(0,0)}$

根据基尼系数的定义, 可计算联合观察与预测反

应的人数比例分布的基尼系数，即 $Gini_{\boldsymbol{\alpha}_l} = f_{lj(1,1)}(1-f_{lj(1,1)})+f_{lj(1,0)}(1-f_{lj(1,0)})+f_{lj(0,1)}(1-f_{lj(0,1)})+f_{lj(0,0)}(1-f_{lj(0,0)})$。$Gini_{\boldsymbol{\alpha}_l}$ 代表 $\boldsymbol{\alpha}_l$ 类被试 4 种反应类别的纯度。其值越小，表明随机抽取的两个 KS 为 $\boldsymbol{\alpha}_l$ 的被试所属不同反应类别的概率越小。在无失误无猜测的条件下，被试 $\boldsymbol{\alpha}_l$ 的理想反应为 1 或 0。若被试 $\boldsymbol{\alpha}_l$ 的理想反应为 1 且 q 向量正确时，则观察反应为 1 的被试中预测反应为 1 的人数比例也越高，即期望 $f_{lj(1,1)}$ 越大，$f_{lj(0,0)}$、$f_{lj(0,1)}$ 和 $f_{lj(1,0)}$ 越小，4 种反应类别的纯度越高；当被试 $\boldsymbol{\alpha}_l$ 的理想反应为 0 且 q 向量正确时，则观察反应为 0 的被试中预测反应为 0 的人数比例也越高，即期望 $f_{lj(0,0)}$ 越大，$f_{lj(1,1)}$、$f_{lj(0,1)}$ 和 $f_{lj(1,0)}$ 越小，4 种反应类别的纯度也越高。因此，正确 q 向量对应的基尼系数应该越小。由表 1 易知 $f_{lj(1,1)}$、$f_{lj(1,0)}$、$f_{lj(0,1)}$ 和 $f_{lj(0,0)}$ 均大于等于 0，且至少有一个大于 0，即 $Gini_{\boldsymbol{\alpha}_l}$ 的值恒大于 0。

实际上，被试总体来自多种 KS，要使得每一种 KS 的 $Gini_{\boldsymbol{\alpha}_l}$ 最小，就等价于要求所有 KS 的基尼系数的和取最小。于是，本研究提出选择使 $\sum_{l=1}^{2^K} Gini_{\boldsymbol{\alpha}_l}$ 最小的 q 向量作为项目 j 的属性模式的方法，称为 ORDP 方法。

该方法适用于所有 CDM。以决定性输入、噪音与门模型(deterministic input, noisy and gate model, DINA) (de la Torre, 2009)为例，对 $\sum_{l=1}^{2^K} Gini_{\boldsymbol{\alpha}_l}$ 的计算过程进行详细说明。

首先，在无失误无猜测的情况下将 KS 分为两类：理想反应为 1 的 KS 类 $\boldsymbol{\alpha}_{lu}(u=1,2,\cdots,U)$ 和理想反应为 0 的 KS 类 $\boldsymbol{\alpha}_{lv}(v=1,2,\cdots,V)$。令 $s_{\boldsymbol{q}_{jc}}$ 与 $g_{\boldsymbol{q}_{jc}}$ 代表项目 j 取第 c 种可能属性模式时的失误和猜测参数。那么，对于理想反应为 1 的被试类 $\boldsymbol{\alpha}_{lu}$ (即 $\prod_{k=1}^{K} \boldsymbol{\alpha}_{luk}^{\boldsymbol{q}_{jkc}} = 1$)，有 $P(y_j=1|\boldsymbol{\alpha}_{lu}) = 1-s_{\boldsymbol{q}_{jc}}$；而对于理想反应为 0 的被试类 $\boldsymbol{\alpha}_{lv}$ (即 $\prod_{k=1}^{K} \boldsymbol{\alpha}_{lvk}^{\boldsymbol{q}_{jkc}} = 0$)，则有 $P(y_j=1|\boldsymbol{\alpha}_{lv}) = g_{\boldsymbol{q}_{jc}}$。

其次，假设知识状态为 $\boldsymbol{\alpha}_{lu}$ 的被试有 N_{lu} 名，其中观察反应为 1 和 0 的人数分别为 r_{lju} 和 $N_{lu}-r_{lju}$。于是，该类被试 4 个反应类别的人数比例分布如表 2 所示。

那么，理想反应为 1 的所有被试类的人数比例分布的基尼系数可化简为

$$\sum_{u=1}^{U} Gini_{\boldsymbol{\alpha}_{lu}} = \sum_{u=1}^{U} (1-(1+2s_{\boldsymbol{q}_{jc}}^2-2s_{\boldsymbol{q}_{jc}})\left(1+2\left(\frac{r_{lju}}{N_{lu}}\right)^2-2\frac{r_{lju}}{N_{lu}}\right) \quad (1)$$

同样，对于知识状态为 $\boldsymbol{\alpha}_{lv}$ 的被试，可令 N_{lv} 和 r_{ljv} 分别代表总人数和答对项目 j 的人数。于是，该类被试 4 个反应类别的人数比例分布如表 3 所示。

那么，理想反应为 0 的所有被试类的人数比例分布的基尼系数为

$$\sum_{v=1}^{V} Gini_{\boldsymbol{\alpha}_{lv}} = \sum_{v=1}^{V} (1-(1+2g_{\boldsymbol{q}_{jc}}^2-2g_{\boldsymbol{q}_{jc}})\left(1+2\left(\frac{r_{ljv}}{N_{lv}}\right)^2-2\frac{r_{ljv}}{N_{lv}}\right) \quad (2)$$

最后，DINA 模型下被试总体在项目 j 上的最优反应分布纯度为 $ORDP = \sum_{u=1}^{U} Gini_{\boldsymbol{\alpha}_{lu}} + \sum_{v=1}^{V} Gini_{\boldsymbol{\alpha}_{lv}}$。

再以实际数据为例说明具体计算过程。为方便计算，假设 $K=2$，则被试的 KS 有 4 种，分别是 $\boldsymbol{\alpha}_1=(1,1)$、$\boldsymbol{\alpha}_2=(1,0)$、$\boldsymbol{\alpha}_3=(0,1)$ 和 $\boldsymbol{\alpha}_4=(0,0)$。假设每一种 KS 均有 100 人，其中观察反应为 1 的人数分别为 r_1=80、r_2=60、r_3=40 和 r_4=20。候选 $\boldsymbol{q}_{jc}$ 有 3 种，分别是 $\boldsymbol{q}_1=(1,1)$、$\boldsymbol{q}_2=(1,0)$ 和 $\boldsymbol{q}_3=(0,1)$。假设所有候选 $\boldsymbol{q}_{jc}$ 的项目参数均为 s=0.1 和 g=0.2。

表 2 被试 $\boldsymbol{\alpha}_{lu}$ 在项目 j 上 4 种反应类别的人数比例分布

反应类别	(*O*1, *E*1)	(*O*1, *E*0)	(*O*0, *E*1)	(*O*0, *E*0)
期望人数比例	$\frac{r_{lju}\cdot(1-s_{\boldsymbol{q}_{jc}})}{N_{lu}}$	$\frac{r_{lju}\cdot s_{\boldsymbol{q}_{jc}}}{N_{lu}}$	$\frac{(N_{lu}-r_{lju})\cdot(1-s_{\boldsymbol{q}_{jc}})}{N_{lu}}$	$\frac{(N_{lu}-r_{lju})\cdot s_{\boldsymbol{q}_{jc}}}{N_{lu}}$

表 3 被试 $\boldsymbol{\alpha}_{lv}$ 在项目 j 上 4 种反应类别的人数比例分布

反应类别	(*O*1, *E*1)	(*O*1, *E*0)	(*O*0, *E*1)	(*O*0, *E*0)
期望人数比例	$\frac{r_{ljv}\cdot g_{\boldsymbol{q}_{jc}}}{N_{lv}}$	$\frac{r_{ljv}\cdot(1-g_{\boldsymbol{q}_{jc}})}{N_{lv}}$	$\frac{(N_{lv}-r_{ljv})\cdot g_{\boldsymbol{q}_{jc}}}{N_{lv}}$	$\frac{(N_{lv}-r_{ljv})\cdot(1-g_{\boldsymbol{q}_{jc}})}{N_{lv}}$

以 $\boldsymbol{\alpha}_1=(1,1)$ 和 $\boldsymbol{\alpha}_2=(1,0)$ 类被试为例。当 $\boldsymbol{q}_1=(1,1)$ 时，$\boldsymbol{\alpha}_1=(1,1)$ 类被试的理想反应为 1，根据表 2 可得 $(O1,E1)$、$(O1,E0)$、$(O0,E1)$ 和 $(O0,E0)$ 这 4 类反应的人数比例依次为 80×(1–0.1))/100=0.72, (80×0.1)/100=0.08, ((100–80)×(1–0.1))/100=0.18 和((100–80)×0.1)/100=0.02。于是 $\boldsymbol{\alpha}_1=(1,1)$ 类被试的基尼系数为 $Gini_{\boldsymbol{\alpha}_1}$ =0.72×(1–0.72)+0.08×(1–0.08)+0.18×(1–0.18)+0.02×(1–0.02)1=0.444。当 $\boldsymbol{q}_1=(1,1)$ 时，$\boldsymbol{\alpha}_2=(1,0)$ 类被试的理想反应为 0，根据表 3 可得这 4 类反应的人数比例分别为(60×0.2)/100=0.12, (60×(1–0.2))/100=0.48, ((100–60)×0.2)/100=0.08 和((100–60)×(1–0.2))/100=0.32。于是 $\boldsymbol{\alpha}_2=(1,0)$ 类被试的基尼系数为 $Gini_{\boldsymbol{\alpha}_2}$ =0.12×(1–0.12)+0.48×(1–0.48)+0.08×(1–0.08)+0.32×(1–0.32)=0.648。

以此类推，可计算出 $Gini_{\boldsymbol{\alpha}_3}$ 和 $Gini_{\boldsymbol{\alpha}_4}$ 的值，将 $Gini_{\boldsymbol{\alpha}_1}$ 到 $Gini_{\boldsymbol{\alpha}_4}$ 的值相加便得到当 $\boldsymbol{q}_1=(1,1)$ 时，被试总体的基尼系数为 $ORDP_{\boldsymbol{q}_1}$ =0.285。同理可得 $\boldsymbol{q}_2=(1,0)$ 和 $\boldsymbol{q}_3=(0,1)$ 时被试总体在项目 j 上的最优反应分布纯度，具体如表 4 所示。

从表 4 可知，$ORDP_{\boldsymbol{q}_3}$ 的值最小，于是项目 j 的正确 q 向量应为 (0,1)。罗芬等人(2020)曾将基尼系数用于双目标 CD-CAT 选题策略。具体而言，他们根据当前已作答的项目和待施测下一个项目的预测反应获得被试 KS 的后验分布，通过使预测的 KS 后验分布的纯度越高即该分布的基尼系数越小为被试选择下一个项目。本研究针对被试在项目 j 上的 4 个反应类别建立了人数比例分布的基尼系数，通过最优该分布的纯度来进行 q 向量的修正，是合理的和可行的。

2.2　基于模型数据拟合的已有 Q 矩阵修正方法

为了考察新方法的表现，研究选择将 ORDP 与 R、RMSEA 和 HD 方法进行比较。原因如下：第一，它们都属于数据绝对拟合指标。其中, ORDP、R、RMSEA、S 统计量和残差方法是基于模型数据拟合视角的绝对拟合指标; HD 方法是基于统计视角的非参数绝对拟合指标。特别地，R、RMSEA 和 HD 方法的计算比较简单。第二，方法间的比较不够。目前，仅 Yu 和 Cheng (2020)比较了 R 和 S 统计量方法。他们的结果表明 R 方法在 DINA 模型下的修正效果优于 S 统计量方法。下面对 R、RMSEA 和 HD 方法依次进行简单介绍。

首先, Yu 和 Cheng (2020)基于观察反应与理想反应的残差 $y_{ij}-\eta_{ij}$ 提出加权的残差统计量指标 R, 见(3)式,

$$R_j=\sum_{i=1}^{N}\log\left[\frac{y_{ij}-\eta_{ij}}{P(y_{ij}\mid\boldsymbol{\alpha}_i)}\right]^2 \tag{3}$$

其中，$P(y_{ij}\mid\boldsymbol{\alpha}_i)$ 表示被试 $\boldsymbol{\alpha}_i$ 在项目 j 上的正确作答概率。

其次，Kang 等人(2019)将近似误差均方根 RMSEA 用于计算被试总体观察作答概率分布与期望作答概率分布的差异，即

$$RMSEA_j=\sqrt{\sum_{z=0}^{1}\sum_{l=1}^{2^K}w(\boldsymbol{\alpha}_l)\left(P(y_j=z\mid\boldsymbol{\alpha}_l)-\frac{n_{jlz}}{N_l}\right)^2} \tag{4}$$

其中，n_{jlz} 表示 N_l 中在项目 j 上得 z 分的人数; $w(\boldsymbol{\alpha}_l)$ 表示总体中 $\boldsymbol{\alpha}_l$ 的后验概率，根据 KS 的先验分布和 $\boldsymbol{\alpha}_l$ 类被试反应的似然计算而得。本文假设 KS 服从均匀分布。

表 4　被试总体在项目 j 不同候选 q 向量下的基尼系数

候选 q 向量	KS	$(O1, E1)$	$(O1, E0)$	$(O0, E1)$	$(O0, E0)$	*Gini*	ORDP
$\boldsymbol{q}_1$=(1,1)	$\boldsymbol{\alpha}_1$=(1,1)	0.202	0.074	0.148	0.020	0.444	2.285
	$\boldsymbol{\alpha}_2$=(1,0)	0.106	0.250	0.074	0.218	0.648	
	$\boldsymbol{\alpha}_3$=(0,1)	0.074	0.218	0.106	0.259	0.657	
	$\boldsymbol{\alpha}_4$=(0,0)	0.038	0.134	0.134	0.230	0.536	
$\boldsymbol{q}_2$=(1,0)	$\boldsymbol{\alpha}_1$=(1,1)	0.202	0.074	0.148	0.020	0.444	2.209
	$\boldsymbol{\alpha}_2$=(1,0)	0.248	0.056	0.230	0.038	0.572	
	$\boldsymbol{\alpha}_3$=(0,1)	0.074	0.218	0.106	0.259	0.657	
	$\boldsymbol{\alpha}_4$=(0,0)	0.038	0.134	0.134	0.230	0.536	
$\boldsymbol{q}_3$=(0,1)	$\boldsymbol{\alpha}_1$=(1,1)	0.202	0.074	0.148	0.020	0.444	2.196
	$\boldsymbol{\alpha}_2$=(1,0)	0.106	0.250	0.074	0.218	0.648	
	$\boldsymbol{\alpha}_3$=(0,1)	0.230	0.038	0.248	0.056	0.572	
	$\boldsymbol{\alpha}_4$=(0,0)	0.038	0.134	0.134	0.230	0.536	

最后，海明距离是一种非参数方法(汪大勋 等, 2018)。它通过最小化全体被试在项目 j 上观察反应向量和理想反应向量的距离 $\sum_{i=1}^{N}|y_{ij}-\eta_{ijc}|$ 来估计 Q 矩阵，可用于修正 Q 矩阵。

2.3 Q 矩阵修正的步骤

令初始 Q 矩阵为 Q^0。Q 矩阵修正的具体算法如下：

第一，对于待修正的项目 j，仅将 Q^0 中项目 j 的初始 q 向量替换为某种可能的属性模式 $\boldsymbol{q}_{jc}$，得到 Q^0_{jc}；

第二，基于 Q^0_{jc} 和作答数据，使用 EM 算法(de la Torre, 2009)估计 $\boldsymbol{q}_{jc}$ 下对应的项目参数 $s_{\boldsymbol{q}_{jc}}$、$g_{\boldsymbol{q}_{jc}}$ 和被试 KS;

第三，计算项目 j 在候选 $\boldsymbol{q}_{jc}$ 下的 ORDP、R、RMSEA 或 HD 的值;

第四，重复步骤一至步骤三，计算项目 j 在所有候选 q 向量下的 ORDP、R、RMSEA 或 HD 的值;

第五，从项目 j 的 C 种可能的属性模式中选择使 ORDP、R、RMSEA 或 HD 最小的 $\boldsymbol{q}_{jc}$ 作为项目 j 的 q 向量;

第六，重复上述步骤，直到修正完所有 L 个项目，算法停止。

3 模拟研究：ORDP、R、RMSEA 和 HD 法在 Q 矩阵修正中的比较

3.1 研究目的

为在复杂测验条件下验证和比较 ORDP、R、RMESA 和 HD 方法在 Q 矩阵修正中的表现，研究考虑了 6 个实验变量。具体包括：两种 KS 分布(均匀分布和多元正态分布)、两种被试人数(N = 300, 1000)、两种测验长度(L = 20, 30)、两种 Q 矩阵错误率(M = 20%, 40%)、两种项目质量(高低 Iq 的参数取值范围分别为[0.05, 0.25]和[0.05, 0.4])和 4 种属性层级结构(独立型，直线型，收敛型和分支型)。属性层级结构见图 1。

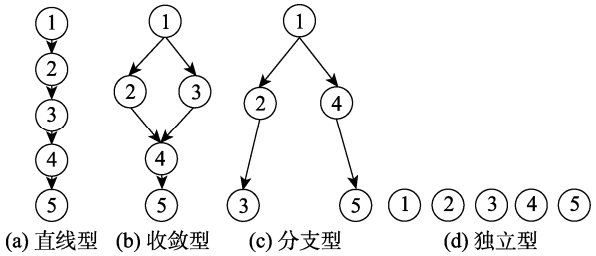

图 1 五个属性的 4 种层级结构图

3.2 数据模拟方法

3.2.1 Q 矩阵的生成

采用蔡艳等人(2013)的方法生成真实 Q 矩阵，即要求测验 Q 矩阵至少包含一个 1 个 R 阵，剩余项目的 q 向量在所有可能的属性考察模式(依属性层级结构的不同而不同)中随机生成。

在真实 Q 矩阵的基础上按项目错误率随机确定相应比例的项目。然后，针对每个项目从所有可能的属性考察模式(本身除外)中随机选择一种作为该项目的错误 q 向量，从而可得错误 Q 矩阵。这种方式产生的错误 q 向量包括了属性冗余、缺失或两者兼有的情况，符合实际情况。事实上，如果 Q 矩阵错误率高达 40%，一般建议重新标定 Q 矩阵。而在模拟实验中，设置高错误率的 Q 矩阵更能考察方法的效能。

3.2.2 被试 KS 和项目参数的生成

无论 KS 服从均匀分布还是多元正态分布，被试均从所有可能 KS 的分布中随机产生。不同的是，各种 KS 的比例因分布的不同而不同。具体而言，当 KS 服从均匀分布(uniform distribution)时，被试总体中各种可能的 KS 的比例相同(Wang, Song, et al., 2018)；当 KS 服从多元正态分布(multidimensional normal distribution)时，首先参考已有研究假设属性间的相关为 0.5 (Chen, 2017; Kang et al., 2019; Wang et al., 2020)，然后根据 Liu 等人(2021)的研究，通过模拟可获得不同层级结构下被试总体中各类 KS 的比例。

DINA 模型的失误 s 和猜测 g 参数均服从均匀分布。其中，高质量项目的 s 和 g 参数从区间(0.05, 0.25)中随机产生，低质量项目的 s 和 g 参数则从区间(0.05, 0.4)中随机产生。

3.2.3 作答反应的生成

基于真实 Q 矩阵、被试 KS 和项目参数，采用 DINA 模型计算被试 i 正确作答项目 j 的概率 P_{ij}，并与随机数 S_{ij} 比较。如果 P_{ij} 大于 S_{ij}，则被试 i 在项目 j 上的反应为 1，否则为 0。

3.3 评价指标

为考察不同方法对 Q 矩阵修正率、项目参数返真率和 KS 估计的影响，采用以下 5 种评价指标：(1) q 向量被完全判准的比例，简称模式判准率(pattern match ratio, PMR); (2) 正确属性被保留的比例(true positive rate, TPR); (3) 错误属性被修改正确的比例(false positive rate, FPR); (4) 基于修正 Q 矩阵的参数估计值 s 和 g 的近似误差均方根的均

值(记为 ME_{sg}); (5) 由修正 Q 矩阵得到的被试 KS 的模式判准率(记为 IMP)。研究使用 R 语言程序，自编计算机代码进行模拟研究，每种实验条件重复 100 次，计算各次实验的均值作为最终结果。PMR 和 ME_{sg} 的计算如下：

$$PMR=\frac{\sum_{t=1}^{100}\sum_{j=1}^{L}I(q_{jt}=q_{jt}^{true})}{L\times 100} \tag{5}$$

$$ME_{sg}=\frac{\sum_{t=1}^{100}\left(\sqrt{\sum_{j=1}^{L}(\hat{s}_{jt}-s_{jt})^2/L}+\sqrt{\sum_{j=1}^{L}(\hat{g}_{jt}-g_{jt})^2/L}\right)}{100\times 2} \tag{6}$$

若修正后项目 j 的 q 向量与真实 q 向量完全一致，则指示函数 $I(q_{jt}=q_{jt}^{true})=1$，否则 $I(q_{jt}=q_{jt}^{true})=0$。PMR、TPR 和 FPR 从不同方面反映 Q 矩阵修正结果，值越高，修正效果越好；ME_{sg} 代表项目参数返真性，值越小，参数估计越好；IMP 表示被试 KS 返真性，值越高，被试诊断分类越准确。

3.4　结果

表 5~8 分别呈现了不同 KS 分布和不同 Q 矩阵错误率时 4 种方法在所有实验条件下的 PMR、TPR 和 FPR 值。表中加粗的数据是相同实验条件下的最优结果。由于各方法修正后项目参数和被试 KS 的返真率差异不大，且与 PMR、TPR 和 FPR 得到的结论一致，为行文简洁，文中未呈现 ME_{sg} 和 IMP 的结果。如有需要，可联系作者。

3.4.1　KS 服从均匀分布时的结果

第一，由表 5 和表 6 可知，ORDP 方法在绝大多数实验条件下都具有最高 PMR 和 TPR 值，在大部分条件下具有最高 FPR 值，接下来依次为 HD、RMSEA 和 R 方法。4 种方法在所有实验条件下的 PMR、TPR 和 FPR 均值从高到低依次为：ORDP (0.916; 0.990; 0.949)、HD (0.914; 0.988; 0.950)、

表 5　KS 服从均匀分布且 M = 20%时各方法在不同实验条件下的 PMR、TPR 和 FPR

评价指标	层级结构	N	高 Iq								低 Iq							
			L = 20				L = 30				L = 20				L = 30			
			ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD
PMR	独立	300	**0.933**	0.830	0.893	0.930	**0.976**	0.908	0.969	0.971	**0.735**	0.590	0.732	0.733	**0.864**	0.738	0.861	0.862
		1000	**0.937**	0.897	0.918	0.935	**0.991**	0.968	0.981	0.983	**0.822**	0.636	0.818	0.821	**0.946**	0.833	0.936	0.941
	直线	300	**0.980**	0.859	0.970	0.974	**0.996**	0.886	0.994	0.994	**0.949**	0.805	0.944	0.946	**0.972**	0.821	0.971	**0.972**
		1000	**0.988**	0.877	0.986	**0.988**	**0.999**	0.900	0.998	**0.999**	**0.967**	0.817	**0.967**	0.966	**0.994**	0.849	0.993	**0.994**
	收敛	300	**0.982**	0.876	0.975	0.977	**0.996**	0.897	0.995	0.995	**0.938**	0.769	0.936	0.937	0.975	0.813	**0.980**	0.973
		1000	**0.990**	0.894	0.983	**0.990**	**0.998**	0.920	0.996	0.997	0.961	0.813	**0.963**	0.961	0.985	0.840	**0.987**	0.984
	分支	300	**0.973**	0.883	0.966	0.969	**0.993**	0.911	0.987	0.991	**0.916**	0.756	0.907	0.911	0.955	0.799	**0.958**	0.955
		1000	**0.990**	0.912	0.984	0.985	**0.997**	0.952	0.996	0.997	**0.950**	0.806	0.949	**0.950**	**0.987**	0.851	0.981	0.984
TPR	独立	300	**0.991**	0.964	0.986	0.988	0.997	0.981	**0.998**	0.994	**0.943**	0.905	0.933	0.940	**0.972**	0.940	0.970	0.971
		1000	**0.994**	0.979	0.992	0.992	**0.999**	0.993	**0.999**	**0.999**	**0.968**	0.918	0.961	0.961	**0.993**	0.966	0.992	0.991
	直线	300	**0.998**	0.974	0.990	0.995	**0.999**	0.979	0.998	0.998	**0.996**	0.960	0.989	0.992	**0.996**	0.962	0.995	0.995
		1000	**0.999**	0.980	**0.999**	**0.999**	**1**	0.983	**1**	**1**	**0.999**	0.966	0.996	0.997	**1**	0.969	**1**	0.999
	收敛	300	**0.998**	0.977	0.996	0.996	0.999	0.981	**1**	0.999	**0.993**	0.956	0.989	0.991	0.996	0.963	**0.997**	0.996
		1000	**0.999**	0.982	0.998	**0.999**	**1**	0.985	**1**	**1**	**0.999**	0.964	0.998	**0.999**	**0.999**	0.968	**0.999**	**0.999**
	分支	300	**0.996**	0.978	**0.996**	**0.996**	**0.999**	0.981	0.998	0.997	**0.988**	0.948	0.984	0.985	0.992	0.960	**0.994**	0.991
		1000	0.998	0.981	**0.999**	0.998	**1**	0.991	**1**	**1**	**0.996**	0.963	0.995	**0.996**	**0.999**	0.970	0.998	**0.999**
FPR	独立	300	0.946	**0.955**	0.916	0.936	**0.981**	**0.981**	0.974	0.979	0.787	**0.847**	0.803	0.790	0.826	**0.910**	0.908	0.825
		1000	0.954	**0.978**	0.946	0.951	**0.993**	**0.993**	0.981	0.992	0.819	**0.893**	0.876	0.818	0.897	0.950	**0.955**	0.891
	直线	300	**0.989**	0.956	0.974	0.979	**0.997**	0.966	0.994	0.996	**0.955**	0.932	0.952	0.950	0.976	0.948	**0.978**	0.976
		1000	**0.995**	0.959	0.988	0.993	**0.999**	0.970	**0.999**	0.998	**0.964**	0.937	**0.964**	0.962	0.994	0.957	**0.995**	0.993
	收敛	300	**0.988**	0.966	0.984	0.983	**0.997**	0.972	0.996	0.995	**0.946**	0.930	**0.946**	**0.946**	0.983	0.945	**0.986**	0.979
		1000	**0.994**	0.968	0.987	0.990	**0.999**	0.981	0.996	0.996	0.958	0.948	**0.960**	0.958	**0.988**	0.959	**0.988**	**0.988**
	分支	300	**0.984**	0.967	0.970	0.973	**0.997**	0.986	0.989	0.994	**0.936**	0.930	0.932	0.933	**0.971**	0.952	**0.971**	0.970
		1000	**0.995**	0.985	0.986	0.988	**0.998**	0.989	**0.998**	0.995	0.953	0.947	**0.954**	0.952	**0.993**	0.970	0.991	0.992

表 6 KS 服从均匀分布且 M = 40%时各方法在不同实验条件下的 PMR、TPR 和 FPR

评价指标	层级结构	N	高 Iq								低 Iq							
			L = 20				L = 30				L = 20				L = 30			
			ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD
PMR	独立	300	**0.751**	0.669	0.668	0.748	**0.887**	0.874	0.859	0.883	**0.574**	0.429	0.537	0.572	**0.685**	0.605	0.675	0.682
		1000	**0.764**	0.708	0.702	0.760	**0.952**	0.922	0.930	0.945	**0.599**	0.452	0.588	0.593	**0.788**	0.706	0.773	0.784
	直线	300	**0.928**	0.787	0.910	0.924	**0.972**	0.836	0.969	0.970	**0.839**	0.733	0.833	0.838	**0.919**	0.761	0.912	0.914
		1000	**0.949**	0.823	0.921	0.939	**0.984**	0.847	0.983	0.980	**0.878**	0.751	0.856	0.871	**0.966**	0.792	0.959	0.961
	收敛	300	**0.923**	0.805	0.884	0.914	**0.968**	0.851	0.961	0.965	**0.830**	0.718	0.809	0.827	**0.902**	0.764	0.885	0.898
		1000	**0.940**	0.827	0.900	0.936	**0.983**	0.880	0.967	0.977	**0.862**	0.739	0.820	0.859	**0.940**	0.805	0.930	0.939
	分支	300	**0.897**	0.806	0.874	0.892	**0.968**	0.885	**0.968**	0.964	**0.799**	0.673	0.781	0.794	**0.895**	0.765	0.893	**0.895**
		1000	**0.927**	0.844	0.887	0.925	**0.986**	0.924	0.972	**0.986**	**0.826**	0.718	0.818	**0.826**	**0.945**	0.809	0.944	0.944
TPR	独立	300	**0.970**	0.931	0.955	0.963	**0.991**	0.976	0.990	**0.991**	**0.909**	0.856	0.904	0.905	**0.948**	0.914	0.945	0.946
		1000	**0.973**	0.939	0.968	0.970	**0.996**	0.985	0.995	0.995	0.927	0.867	**0.930**	0.919	**0.972**	0.937	0.971	0.971
	直线	300	**0.995**	0.967	0.984	0.994	**0.998**	0.973	0.993	0.996	**0.990**	0.956	0.979	0.985	**0.993**	0.954	0.990	0.991
		1000	**0.997**	0.973	0.990	0.995	**0.999**	0.976	**0.999**	**0.999**	**0.993**	0.958	0.985	0.991	**0.998**	0.962	**0.998**	**0.998**
	收敛	300	**0.993**	0.969	0.982	0.984	**0.997**	0.976	0.995	0.994	**0.985**	0.952	0.969	0.980	**0.991**	0.958	0.985	0.990
		1000	**0.994**	0.974	0.991	0.992	**0.999**	0.982	**0.999**	**0.999**	**0.991**	0.957	0.980	0.988	**0.998**	0.968	0.995	0.995
	分支	300	**0.993**	0.967	0.987	**0.993**	0.997	0.979	**0.999**	0.993	**0.980**	0.936	0.977	**0.980**	**0.991**	0.956	**0.991**	0.990
		1000	**0.994**	0.972	0.992	0.993	**0.999**	0.987	**0.999**	0.995	**0.989**	0.948	0.983	0.985	**0.997**	0.966	0.996	0.996
FPR	独立	300	0.854	**0.880**	0.820	0.850	0.939	**0.962**	0.923	0.934	0.728	**0.766**	0.731	0.723	0.810	**0.850**	0.818	0.812
		1000	0.862	**0.901**	0.838	0.861	0.971	**0.977**	0.960	0.965	0.754	**0.782**	0.761	0.753	0.882	**0.906**	0.874	0.876
	直线	300	**0.970**	0.933	0.948	0.962	**0.988**	0.956	0.985	0.983	0.910	**0.911**	0.899	0.908	**0.957**	0.929	0.950	0.955
		1000	**0.978**	0.950	0.957	0.973	**0.994**	0.960	0.991	**0.994**	**0.933**	0.919	0.912	0.929	**0.983**	0.940	0.977	0.979
	收敛	300	**0.960**	0.942	0.946	0.952	**0.987**	0.961	0.982	0.982	**0.905**	0.904	0.895	0.896	**0.949**	0.930	0.942	**0.949**
		1000	**0.973**	0.947	0.948	0.971	**0.993**	0.965	0.983	0.990	**0.931**	0.922	0.906	0.922	**0.966**	0.943	0.962	0.965
	分支	300	**0.947**	0.940	0.932	0.940	**0.987**	0.971	0.982	0.982	0.885	**0.898**	0.875	0.885	**0.944**	0.935	0.943	0.942
		1000	**0.964**	0.956	0.938	0.960	**0.993**	0.987	0.985	0.989	0.903	**0.917**	0.892	0.896	**0.969**	0.950	**0.969**	0.968

表 7 KS 服从多元正态分布且 M = 20%时各方法在不同实验条件下的 PMR、TPR 和 FPR

评价指标	层级结构	N	高 Iq								低 Iq							
			L = 20				L = 30				L = 20				L = 30			
			ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD
PMR	独立	300	**0.909**	0.643	0.874	0.894	**0.961**	0.703	0.941	0.956	**0.724**	0.491	0.713	0.716	**0.843**	0.615	0.795	0.836
		1000	**0.951**	0.763	0.910	0.948	**0.987**	0.815	0.966	0.980	**0.819**	0.552	0.811	0.814	**0.937**	0.737	0.930	0.931
	直线	300	0.918	0.613	**0.923**	0.913	0.978	0.635	**0.985**	0.974	0.860	0.570	**0.869**	0.858	0.917	0.616	**0.925**	0.915
		1000	0.973	0.643	**0.978**	0.965	0.983	0.649	**0.991**	0.980	0.902	0.644	**0.908**	0.901	0.959	0.636	**0.968**	0.956
	收敛	300	0.918	0.608	**0.927**	0.919	0.955	0.639	**0.965**	0.953	0.863	0.590	**0.870**	0.860	0.898	0.626	**0.905**	0.896
		1000	0.926	0.638	**0.931**	0.924	0.973	0.681	**0.982**	0.974	0.885	0.633	**0.891**	0.882	0.948	0.665	**0.953**	0.942
	分支	300	0.930	0.791	**0.939**	0.927	0.969	0.816	**0.976**	0.968	0.854	0.689	**0.860**	0.848	0.869	0.715	**0.875**	0.866
		1000	0.948	0.824	**0.952**	0.944	0.980	0.867	**0.989**	0.978	0.900	0.767	**0.906**	0.901	0.928	0.799	**0.935**	0.920
TPR	独立	300	**0.992**	0.928	0.985	0.984	**0.998**	0.941	0.994	0.994	**0.952**	0.872	0.940	0.948	**0.965**	0.915	0.955	0.960
		1000	**0.996**	0.954	0.995	0.995	**1**	0.965	0.999	0.997	**0.983**	0.900	0.971	0.979	**0.992**	0.943	0.984	0.988
	直线	300	**0.990**	0.898	0.980	0.984	0.996	0.904	**0.999**	0.996	0.977	0.862	**0.980**	0.973	0.984	0.875	**0.990**	0.982
		1000	0.998	0.922	**0.999**	0.995	0.998	0.916	**1**	0.998	0.992	0.888	**0.996**	0.990	0.991	0.896	**0.998**	0.993
	收敛	300	0.986	0.904	**0.987**	0.985	0.991	0.912	**0.995**	0.990	0.975	0.890	**0.982**	0.974	0.979	0.902	**0.988**	0.976
		1000	**0.991**	0.927	0.988	0.988	0.997	0.928	**1**	0.998	0.984	0.910	**0.991**	0.982	0.989	0.917	**0.997**	0.985

续表

评价指标	层级结构	N	高 Iq								低 Iq							
			L = 20				L = 30				L = 20				L = 30			
			ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD
FPR	分支	300	0.990	0.960	**0.994**	0.990	0.990	0.965	**0.996**	0.991	0.962	0.939	**0.977**	0.959	0.976	0.941	**0.980**	0.968
		1000	0.994	0.968	**0.996**	0.992	0.998	0.976	**1**	0.998	0.982	0.955	**0.991**	0.987	0.989	0.959	**0.994**	0.982
	独立	300	**0.932**	0.872	0.912	0.925	**0.973**	0.912	0.962	0.968	**0.766**	0.753	0.748	0.756	**0.878**	0.841	0.870	0.876
		1000	**0.954**	0.925	0.922	0.946	**0.991**	0.948	0.971	0.988	**0.839**	0.814	0.818	0.833	**0.965**	0.912	0.946	0.960
	直线	300	0.945	0.855	**0.950**	0.942	0.985	0.874	**0.989**	0.986	0.879	0.824	**0.888**	0.871	0.935	0.844	**0.941**	0.932
		1000	0.977	0.871	**0.981**	0.976	0.988	0.880	**0.991**	0.987	0.901	0.843	**0.909**	0.900	0.965	0.859	**0.974**	0.964
	收敛	300	0.951	0.878	**0.955**	0.947	0.965	0.897	**0.978**	0.967	0.876	0.821	**0.883**	0.874	0.930	0.861	**0.936**	0.925
		1000	0.953	0.880	**0.961**	0.955	0.977	0.908	**0.982**	0.976	0.899	0.849	**0.906**	0.895	0.949	0.863	**0.956**	0.936
	分支	300	0.935	0.937	**0.941**	0.933	0.979	0.947	**0.988**	0.978	0.888	0.878	**0.896**	0.883	**0.925**	0.911	**0.939**	0.917
		1000	0.966	0.950	**0.972**	0.958	0.983	0.961	**0.990**	0.980	0.920	0.922	**0.927**	0.915	0.947	0.941	**0.953**	0.937

表 8　KS 服从多元正态分布且 M = 40%时各方法在不同实验条件下的 PMR、TPR 和 FPR

评价指标	层级结构	N	高 Iq								低 Iq							
			L = 20				L = 30				L = 20				L = 30			
			ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD	ORDP	R	RMSEA	HD
PMR	独立	300	**0.718**	0.527	0.670	0.713	**0.874**	0.641	0.845	0.870	**0.561**	0.352	0.525	0.556	**0.691**	0.501	0.668	0.689
		1000	**0.766**	0.586	0.733	0.754	**0.923**	0.729	0.880	0.915	**0.609**	0.390	0.587	0.607	**0.785**	0.608	0.765	0.778
	直线	300	0.829	0.568	**0.834**	0.828	0.939	0.603	**0.947**	0.931	0.709	0.519	**0.718**	0.700	0.848	0.569	**0.853**	0.840
		1000	0.864	0.587	**0.873**	0.860	0.957	0.625	**0.964**	0.954	0.776	0.540	**0.783**	0.774	0.866	0.608	**0.873**	0.859
	收敛	300	0.844	0.543	**0.853**	0.839	0.917	0.583	**0.924**	0.911	0.710	0.525	**0.717**	0.709	0.840	0.556	**0.845**	0.837
		1000	0.859	0.586	**0.863**	0.859	0.950	0.613	**0.958**	0.950	0.763	0.541	**0.772**	0.755	0.855	0.600	**0.860**	0.851
	分支	300	0.811	0.706	**0.816**	0.807	0.934	0.809	**0.943**	0.930	0.700	0.602	**0.709**	0.693	0.776	0.673	**0.783**	0.775
		1000	0.840	0.754	**0.848**	0.839	0.953	0.825	**0.961**	0.948	0.740	0.630	**0.746**	0.735	0.840	0.719	**0.848**	0.838
TPR	独立	300	**0.974**	0.898	0.960	0.970	**0.985**	0.929	0.980	0.981	**0.932**	0.842	0.927	0.930	**0.971**	0.888	0.958	0.966
		1000	**0.981**	0.917	0.978	**0.981**	**0.994**	0.946	0.991	**0.994**	**0.943**	0.854	0.932	0.937	**0.989**	0.922	0.980	0.983
	直线	300	**0.984**	0.892	0.974	0.980	0.989	0.898	**0.998**	0.985	0.966	0.864	**0.975**	0.962	0.979	0.891	**0.988**	0.973
		1000	0.987	0.911	**0.989**	0.985	0.996	0.917	**1**	0.995	**0.983**	0.891	**0.983**	0.980	0.989	0.898	**0.994**	0.984
	收敛	300	0.984	0.904	**0.989**	0.984	0.989	0.905	**0.994**	0.986	0.959	0.881	**0.967**	0.954	0.980	0.896	**0.985**	0.974
		1000	**0.991**	0.917	0.990	0.986	0.994	0.920	**0.999**	0.990	0.981	0.900	**0.989**	0.979	0.985	0.902	**0.991**	0.985
	分支	300	0.983	0.950	**0.989**	0.982	0.989	0.968	**0.995**	0.988	0.964	0.933	**0.971**	0.964	0.969	0.941	**0.976**	0.965
		1000	0.989	0.959	**0.991**	0.985	0.996	0.970	**0.999**	0.991	0.971	0.939	**0.980**	0.967	0.980	0.956	**0.989**	0.975
FPR	独立	300	**0.834**	0.816	0.808	0.829	**0.945**	0.893	0.925	0.941	**0.717**	0.688	0.692	0.710	**0.804**	0.784	0.801	0.800
		1000	**0.878**	0.843	0.842	0.870	**0.968**	0.918	0.936	0.962	**0.738**	0.700	0.720	0.731	**0.871**	0.838	0.857	0.865
	直线	300	0.920	0.841	**0.922**	0.913	0.968	0.855	**0.973**	0.965	0.850	0.796	**0.858**	0.849	0.918	0.834	**0.927**	0.911
		1000	0.925	0.844	**0.938**	0.922	0.974	0.875	**0.981**	0.971	0.862	0.801	**0.864**	0.860	0.935	0.839	**0.941**	0.933
	收敛	300	0.925	0.843	**0.930**	0.920	0.955	0.870	**0.963**	0.947	0.853	0.820	0.857	0.853	0.911	0.850	**0.917**	0.910
		1000	0.932	0.862	**0.935**	0.925	0.969	0.887	**0.978**	0.966	0.863	0.831	**0.866**	0.861	0.915	0.860	**0.920**	0.914
	分支	300	0.900	0.888	**0.906**	0.894	0.967	0.948	**0.976**	0.960	0.820	0.818	**0.825**	0.820	0.889	0.880	**0.893**	0.881
		1000	0.917	0.894	**0.919**	0.915	0.975	0.955	**0.980**	0.969	0.855	0.842	**0.863**	0.849	0.913	0.903	**0.922**	0.913

RMSEA (0.904; 0.986; 0.942)和 R (0.803; 0.962; 0.939)。由此可知, ORDP 方法在 Q 矩阵修正中的表现明显优于 RMSEA 和 R 方法，略微优于 HD 方法。另外，各方法的 TPR 均高于 FPR，且 4 种方法对于正确属性的保留率均在 95%以上。换言之，它们对正确属性方面的保留率大于对错误属性的修正率。另外，ORDP 方法对于错误属性的修正能力在低质量项目中更易受属性层级结构的影响。

第二，当仅变化 N、L、M 或 Iq 时可知：(1) N 越大、L 越长、M 越低或 Iq 越高时，PMR、TPR、FPR 和 IMP 的值都越大，ME_{sg} 的值越小，Q 矩阵修正效果越好；(2) 固定 N = 300 (1000)时，ORDP 方法在所有实验条件下 PMR 均值为 0.905 (0.930)。此时，PMR 均值的全距为 0.025。同理，固定 N = 300 (1000)时，R、RMSEA 和 HD 方法在所有实验条件下 PMR 均值的全距分别为 0.038、0.029 和 0.029。类似的，仅固定 L、M 或 Iq 时，各方法在所有实验条件下 PMR 均值的全距分别记为 ORDP (0.065; 0.081; 0.073)、R (0.075; 0.075; 0.122)、RMSEA (0.078; 0.098; 0.064)、HD (0.065; 0.083; 0.071)。总体上，R 和 RMSEA 方法受各因素影响的波动最大，ORDP 方法的波动最小，HD 方法的波动范围居中。

第三，属性层级结构对各方法修正率的影响。在独立、直线、收敛和分支型结构下，4 种方法的 PMR 均值分别为：ORDP (0.825, 0.955, 0.948, 0.938)、RMSEA (0.803, 0.948, 0.936, 0.929)、R (0.735, 0.821, 0.826, 0.831)、HD (0.821, 0.952, 0.946, 0.936)。可见，ORDP、RMSEA 和 HD 方法在各层级结构下表现的优劣依次为直线、收敛、分支和独立型；R 方法除独立型结构下的结构最差外，其它三种层级结构下的结果无明显差异。

3.4.2 KS 服从多元正态分布时的结果

第一，由表 7 和表 8 可知，总体上 RMSEA 方法的 PMR、TPR 和 FPR 值均最高，接下来依次为 ORDP、HD 和 R 方法。4 种方法在所有实验条件下的 PMR、TPR 和 FPR 均值从高到低依次为：RMSEA (0.874; 0.985; 0.919)、ORDP (0.864; 0.983; 0.915)、HD (0.856; 0.980; 0.911)和 R (0.639; 0.918; 0.865)。由此可知，当 KS 服从多元正态分布时，RMSEA 方法的表现整体稍优于 ORDP 方法，平均 PMR 差距在 0.01 以内。此外，所有方法在多元正态分布下的 PMR 值均低于它们在均匀分布下的结果，这与已有研究(Chiu, 2013; Wang et al., 2020; Wang, Song, et al., 2018)的结果一致。

第二，在不同 N、L、M 或 Iq 条件下：(1) 所有方法的 PMR、TPR 和 FPR 均随着 N、L 的增加、M 的降低或 Iq 的提高而增大；(2) 固定 N、L、M 或 Iq 时，各方法在所有实验条件下 PMR 均值的全距分别记为 ORDP (0.040; 0.083; 0.104; 0.096)、R (0.050; 0.064; 0.076; 0.073)、RMSEA (0.042; 0.084; 0.106; 0.093)、HD (0.044; 0.086; 0.093; 0.106)。总体上，HD 方法受 L 和 Iq 的影响较大，而 ORDP 和 RMSEA 方法则更易受 M 的影响。

第三，4 种方法在独立、直线、收敛和分支型结构下的 PMR 均值分别为：ORDP (0.816; 0.892; 0.881; 0.873)、RMSEA (0.788; 0.899; 0.889; 0.880)、R (0.603; 0.602; 0.602; 0.749)、HD (0.810; 0.879; 0.870; 0.866)。可见，当 KS 服从多元正态分布时，独立属性结构中 ORDP 的表现明显优于 RMSEA 方法，其它属性结构中 RMSEA 方法的表现稍优于 ORDP 方法，R 方法仍在所有结构下表现最差。此外，ORDP、RMSEA 和 HD 方法在各层级结构下表现的优劣仍然为直线、收敛、分支和独立型，这与 KS 服从均匀分布时的结果一致。

4 四种方法在分数减法数据 Q 矩阵修正中的应用

基于 Tatsuoka (1984)分数减法数据，研究二运用 4 种方法对专家标定的 Q 矩阵进行修正。该测验包括 15 个项目，考察 5 个属性，一共有 536 名被试的作答反应。初始 Q 矩阵如表 9 中的 0、1 所示。此外，通过原始 Q 矩阵和各方法修正后 Q 矩阵的相对拟合指标和绝对拟合指标比较不同 Q 矩阵的模型数据拟合度。其中，相对拟合指标包括偏差(-2LogLikelihood, -2LL)、赤池信息准则(Akaike information criterion, AIC)和贝叶斯信息准则(Bayesian information criterion, BIC)，绝对拟合指标包括 M_2、RMSEA 和标准均方根残差(standardized root mean square residual, SRMSR)统计量。

表 9~10 分别是各种方法对专家界定的 Q 矩阵的修正情况和模型数据拟合结果。由表 9 可知：ORDP、R、RMSEA 和 HD 方法分别调整了 24、32、5 和 1 个属性。ORDP 方法未调整第 1、3、5、8、9 和 11 题。由表 10 可知，只有 ORDP 方法修正后的 Q 矩阵的相对拟合指标均优于原始 Q 矩阵的值。所有方法修正后的绝对拟合指标均低于原始 Q 矩阵的结果，且 ORDP 方法的 M_2 和 RMSEA 值最低。这表明采用 ORDP 方法修正后的 Q 矩阵与模型的拟合度更优。值得注意的是，各方法提出的修正方案应作为专家修正 Q 矩阵时的建议，研究者不能完全依赖数据分析，而忽视对项目特征的分析。

5 结论与讨论

5.1 结论

Q 矩阵是认知诊断的重要组成部分。它通常由领域专家进行标定，具有一定主观性。因此，开发

表 9　Tatsuoka 分数减法数据的测验 Q 矩阵以及各方法对属性的修正情况

Item	A1	A2	A3	A4	A5	Item	A1	A2	A3	A4	A5
1	1	0^	0*^~	0~	0	9	1	0	1*	0	0
2	1#*	1#*	1#*	1	0	10	1#*	0	1#*	1#	1*
3	1*	0*	0^	0	0	11	1*	0	1	0	0
4	1#*	1#*	1#*	1#*	1	12	1#*	0	1#*	1	0
5	0^	0	1	0	0	13	1#*	1#*	1#*	1	0
6	1#*	1	1#*	1	0	14	1#*	1	1#*	1#*	1
7	1#*	1*	1#*	1	0	15	1#*	1*	1#*	1	0
8	1	1*	0^	0	0						

注：A1~A5 分别表示：运算基础、化简(代)分数、从分数中分离出整数、借位和化整数为分数。“#”、“*”、“^”和“~”分别表示 ORDP、R、RMSEA 和 HD 方法调整的属性。

表 10　基于 4 种方法修正后 Q 矩阵的拟合指标

Q 矩阵	相对拟合指标			绝对拟合指标				
	-2LL	AIC	BIC	M_2			RMSEA	SRMSR
				M_2	df	p		
Q_original	6911.549	7033.550	7294.880	235.320	59	0.001	0.075	0.113
Q_ORDP	**6844.310**	**6966.310**	**7227.640**	**178.526**	59	0.001	**0.062**	0.094
Q_R	6974.382	7096.380	7357.710	179.088	59	0.001	**0.062**	0.093
Q_RMSEA	6932.032	7054.030	7315.360	214.976	59	0.001	0.070	0.093
Q_HD	6904.563	7026.560	7287.890	196.354	59	0.001	0.066	**0.090**

简单高效的 Q 矩阵修正方法是认知诊断的重要研究议题，具有重要实践价值。本研究借鉴基尼系数的定义，构造了预测人数比例分布的基尼系数指标，并通过 Monte Carlo 模拟实验和基于 Tatsuoka (1984) 的分数减法数据，验证和比较了新方法与 R、RMSEA 和 HD 方法在 Q 矩阵修正中的表现。研究表明：

第一，对项目 q 向量的模式判准率、正确属性的保留率和错误属性的修正率而言，当 KS 服从均匀分布时，ORDP 方法整体上最优，接下来依次是 HD、RMSEA 和 R 方法。当 KS 服从多元正态分布时，ORDP 方法在独立层级结构下最优，其它层级结构下 RMSEA 方法稍优于 ORDP 方法；第二，各方法在 KS 服从多元正态分布下的修正效果低于服从均匀分布时的结果；第三，被试人数、测验长度、Q 矩阵错误率、项目质量和属性层级结构对 4 种方法 Q 矩阵修正效果均有明显影响。一般地，人数越少、测验越长、Q 矩阵错误率越高或项目质量越低，各方法的表现越差。其中，ORDP 方法受被试人数影响较小，在小样本条件下仍有较高的修正率；第四，基于实证数据的研究结果表明，ORDP 方法修正后的 Q 矩阵与数据的拟合度最高。

5.2　讨论

研究基于模型数据拟合的角度比较了 4 种 Q 矩阵修正方法。其中，HD 和 R 方法反映了观察反应分布和理想反应分布的差异；RMSEA 方法描述了观察反应概率分布和预测反应概率分布的差异；ORDP 方法则刻画了每类被试依据观察反应获得的预测人数比例分布的一致性。除 HD 方法不运用 CDM 属于非参数数据拟合方法外，其它三种方法都是基于 CDM 的模型数据拟合方法。

模拟研究发现，ORDP 和 RMSEA 方法表现的优劣会因知识状态的不同而不同。当被试为均匀分布时，除部分项目质量低、测验较长的条件外，ORDP 方法的表现均明显优于 RMSEA 方法。而当被试为多元正态分布时，ORDP 方法只有在独立型结构下的修正结果优于 RMSEA 方法。已有研究也表明，不同 Q 矩阵修正方法的优劣会随知识状态分布的不同而改变(Kang et al., 2019; Wang et al., 2020; Wang, Song, et al., 2018)。

模拟研究还借鉴 Kang 等人(2019)的方法采用一次修正的方式。比较 Yu 和 Cheng (2020)和本研究中 R 方法的结果，可知相同条件下采用循环修正和一次修正的结果没有太大差异。事实上，虽然循环修正得到的结果更稳定、更稳健，但是循环修正非常费时。以 ORDP 方法为例，在属性层级结构为独立型，被试知识状态为均匀分布，L = 20, N = 300,

M = 20%, Iq~U [0.05, 0.25]的条件下，循环修正需要 147 s，而一次修正仅需要 12 s。此外，循环修正可能存在前后两次修正的 Q 矩阵始终不相同即不收敛的情况(汪大勋 等, 2019)。一次修正虽然能保证方法之间比较的基础相同且花费时间短，但今后还有待深入比较两种修正方式的差异。

另外，本研究仅基于 DINA 模型开展实验，今后有必要基于其它认知诊断模型考察 ORDP 方法的表现。一般地，项目参数和知识状态的估计精度与 Q 矩阵估计精度相互关联。于是，探讨如何校准项目参数和知识状态的估计误差对提高 Q 矩阵估计(修正)率具有重要意义。此外，随着考试形式和评价方式的多样化，单一的测验条件已不能适应测验需要。因此，未来研究有必要针对多级评分(杭丹丹, 2020; 刘芯伶, 2020; Ma & de la Torre, 2020; 汪大勋 等, 2020)、多解题策略或属性多级等复杂测验条件研究 Q 矩阵估计(修正)方法。最后，探索如何将 Q 矩阵估计(修正)方法运用于在线标定中，以及联合标定 Q 矩阵和项目参数(陈平, 辛涛, 2011; Chen et al., 2015; 谭青蓉, 2019)，都是今后研究的重要方向。

总体上，Q 矩阵估计(修正)方法通过数据分析获得 Q 矩阵，是一种量化研究，随机误差和方法均会影响结果。专家分析作为一种质性研究，易受主观因素的影响。因此，今后一方面可以将专家标定和 Q 矩阵估计(修正)方法相结合，另一方面可以先使用 Q 矩阵估计(修正)方法，将有争议的题目交于专家讨论，从而减少工作量并提高 Q 矩阵估计(修正)方法的可信度。

参 考 文 献

Cai, Y., Tu, D. B., & Ding, S. L. (2013). A simulation study to compare five cognitive diagnostic models. *Acta Psychologica Sinica, 45*(11), 1295–1304.

[蔡艳, 涂冬波, 丁树良. (2013). 五大认知诊断模型的诊断正确率比较及其影响因素: 基于分布形态、属性数及样本容量的比较. *心理学报, 45*(11), 1295–1304.]

Chen, J. S. (2017). A residual-based approach to validate Q-matrix specifications. *Applied Psychological Measurement, 41*(4), 277–293.

Chen, P., & Xin, T. (2011). Item replenishing in cognitive diagnostic computerized adaptive testing. *Acta Psychologica Sinica, 43*(7), 836–850.

[陈平, 辛涛. (2011). 认知诊断计算机化自适应测验中的项目增补. *心理学报, 43*(7), 836–850.]

Chen, Y. H., Culpepper, S. A., Chen, Y. G., & Douglas, J. (2018). Bayesian estimation of the DINA Q-matrix. *Psychometrika, 83*(1), 89–108.

Chen, Y. H., Liu, J., & Ying, Z. (2015). Online item calibration for Q-matrix in CD-CAT. *Applied Psychological Measurement, 39*(1), 5–15.

Chiu, C. Y. (2013). Statistical refinement of the Q-Matrix in cognitive diagnosis. *Applied Psychological Measurement, 37*(8), 598–618.

Chung, M. T. (2019). A Gibbs sampling algorithm that estimates the Q-matrix for the DINA model. *Journal of Mathematical Psychology, 93*(), 102–275.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement, 35*(1), 8–26.

DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement, 36*(6), 447–468.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343–362.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115–130.

de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika, 81*(2), 253–273.

Hang, D. D. (2020). *Research on the validation method and application of Q-matrix in polytomously scored cognitive diagnosis assessment* (Unpublished Master's thesis). Jiangxi Normal University, Nanchang, China.

[杭丹丹. (2020). *多级计分认知诊断评估中的 Q 矩阵验证方法与应用研究* (硕士学位论文). 江西师范大学, 南昌.]

Kang, C. H., Yang, Y. K., & Zeng, P. H. (2019). Q-matrix refinement based on item fit statistic RMSEA. *Applied Psychological Measurement, 43*(7), 527–542.

Liu, J. C., Xu, G. J., & Ying, Z. L. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement, 36*(7), 548–564.

Liu, X. L. (2020). *A nonparametric approach of Q-matrix modification for polytomous situations* (Unpublished Master's thesis). Zhejiang Normal University, Jinhua, China.

[刘芯伶. (2020). *多级评分情境下 Q 矩阵修正的非参数方法* (硕士学位论文). 浙江师范大学, 金华.]

Liu, Y. L., Xin, T., & Jiang, Y. (2021). Structural parameter standard error estimation method in diagnostic classification models: Estimation and application. *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2021.1919048

Luo, F., Wang, X. Q., Cai, Y., Tu, D. B. (2020). A new dual-objective CD-CAT item selection method based on the Gini index. *Acta Psychologica Sinica, 52*(12), 1452–1465.

[罗芬, 王晓庆, 蔡艳, 涂冬波. (2020). 基于基尼指数的双目标 CD-CAT 选题策略. *心理学报, 52*(12), 1452–1465.]

Ma, W. C., & de la Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *The British Journal of Mathematical and Statistical Psychology, 73*(1), 142–163.

Rupp, A. A., & Templin, J. L. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68*(1), 78–96.

Tan, Q. R. (2019). *The development of generalized online calibration methods in CD-CAT* (Unpublished Master's thesis). Jiangxi Normal University, Nanchang, China.

[谭青蓉. (2019). *CD-CAT 广义在线标定方法开发研究* (硕士学位论文). 江西师范大学, 南昌.]

Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Urbana, IL: Computer-based

Education Research Laboratory, University of Illinois.
Tu, D. B., Cai, Y., & Dai, H. Q. (2012). A new method of Q-matrix validation based on DINA model. *Acta Psychologica Sinica*, *44*(4), 558–568.
[涂冬波, 蔡艳, 戴海琦. (2012). 基于 DINA 模型的 *Q* 矩阵修正方法. *心理学报*, *44*(4), 558–568.]
Wang, D. X., Cai, Y. & Tu, D. B. (2020). Q-matrix estimation methods for cognitive diagnosis models: Based on partial known Q-matrix. *Multivariate Behavioral Research*, *56*(3), 514–526.
Wang, D. X., Gao, X. L., Cai, Y., & Tu, D. B. (2019). A new general method for Q-matrix validation in cognitive diagnosis assessments. *Journal of Psychological Science*, *42*(4), 988–996.
[汪大勋, 高旭亮, 蔡艳, 涂冬波. (2019). 一种广义的认知诊断 *Q* 矩阵修正新方法. *心理科学*, *42*(4), 988–996.]
Wang, D. X., Gao, X. L., Cai, Y. & Tu, D. B. (2020). A method of Q-matrix validation for polytomous response cognitive diagnosis model based on relative fit statistics. *Acta Psychologica Sinica*, *52*(1), 93–106.
[汪大勋, 高旭亮, 蔡艳, 涂冬波. (2020). 基于类别水平的多级计分认知诊断 *Q* 矩阵修正：相对拟合统计量视角. *心理学报*, *52*(1), 93–106.]
Wang, D. X., Gao, X. L., Han, Y. T., & Tu, D. B. (2018). A simple and effective Q-matrix estimation method: From non-parametric perspective. *Journal of Psychological Science*, *41*(1), 180–188.
[汪大勋, 高旭亮, 韩雨婷, 涂冬波. (2018). 一种简单有效的 *Q* 矩阵估计方法开发: 基于非参数化方法视角. *心理科学*, *41*(1), 180–188.]
Wang, W. Y., Song, L. H., Ding, S. L., Meng, Y. R., Cao, C. X., & Jie, Y. J. (2018). An EM-based method for Q-matrix validation. *Applied Psychological Measurement*, *42*(6), 446–459.
Yu, X. F., & Cheng, Y. (2020). Data-driven Q-matrix validation using a residual-based statistic in cognitive diagnostic assessment. *British Journal of Mathematical and Statistical Psychology*, *73*, 145–179.

A simple and effective new method of Q-matrix validation

LI Jia, MAO Xiuzhen, WEI Jia

(*Institute of Educational Sichuan Normal University, Chengdu* 610066, *China*)

Abstract

Cognitive diagnostic theory (CDT) can provide fine-grained and multidimensional process assessment results, which has important research and practical values. The Q-matrix that represents the relationship between items and attributes, is the basis of CDT. The accuracy of the Q-matrix is an important factor that affects the accuracy of items parameter estimation and participants' diagnosis. Therefore, it is of great significance to check the correctness of the Q-matrix or to validate it. A lot of studies have been carried out on the estimation or validation of Q-matrix, and a variety of methods have been proposed from different perspectives, each having their advantages and disadvantages. The methods based on model-data fit can provide rich test information without the need of complex parameter estimation and time-consuming and tedious calculation. Following this line of thinking, this study used Gini coefficient to express the purity of expected numbers proportion distribution, and constructed a simple and efficient Q-matrix validation method, called the optimization of response distribution purity (ORDP) method, which is suitable for both simplified model and saturated model.

Residual index (R), root mean square error approximate (RMSEA) and hamming distance (HD) were compared to evaluate the performances with varied influencing factors, under the conditions of two different distribution of knowledge states (KS) (uniform distribution, multidimensional normal distribution), two different sample sizes (300, 1000), two different test lengths (20, 30), Q-matrix error rates (20%, 40%), item qualities ([0.05, 0.25], [0.05, 0.24]) and attribute hierarchical structures (independent structure, linear structure, convergent structure, and branched structure). The specific algorithm of Q-matrix validation is as follows. Firstly, the initial Q-matrix is represented by Q^0. When validating the first item j, the initial q-vector of item j in Q^0 is replaced with one of all possible q-vectors, leaving the rest of the items intact. Then, the EM algorithm is used to estimate the item parameters and the knowledge states of the participants. Lastly, the q-vector that minimizes ORDP, R, RMSEA, or HD for the q-vector of the item is selected.

Simulation results demonstrate that: (1) The distribution of KS affects the performance of each method. Specifically, when the KS is uniformly distributed, ORDP method is superior to other methods, HD method is the next, followed by RMSEA and R methods; When the KS follows multivariate normal distribution, there is no significant difference between RMSEA and ORDP. RMSEA method is slightly better than ORDP method except

independent structure, followed by HD and R method; (2) The validation effect of these methods under multivariate normal distribution is not as good as that under uniform distribution; (3) The validation rates of the four methods all affected by sample sizes, test lengths, Q-matrix error rates, item qualities and attribute hierarchical structures. If the smaller the number of respondents, the shorter the test length, the higher the Q-matrix error rates, or the lower the item quality, the worse the performance of each method will be, and vice versa; (4) The validation results based on the fractional subtraction data of Tatsuoka (1984) show that the Q-matrix modified by ORDP method has the best model-data fit.

In this study, the ORDP index representing the purity of the expected numbers proportion distribution was constructed based on the Gini coefficient. Simulation and empirical studies show that this method has a high validation rate for Q-matrices under different conditions. On the whole, the new method proposed in this study validates the Q-matrix through data analysis, which can reduce the workload of experts and thus improve the correctness of the Q-matrix.

Key words cognitive diagnosis, Q-matrix validation methods, ORDP method, DINA model